Int J Appl Earth Obs Geoinformation 79 (2019) 48–57

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag

Construction of a drought monitoring model using deep learning based on multi-source remote sensing data

Runping Shen[a,*], Anqi Huang[a], Bolun Li[a], Jia Guo[b]

[a] School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China
[b] Inner Mongolia autonomous region Wuhai Meteorological Bureau, Wuhai 016000, China

ARTICLE INFO

Keywords:
Drought
Remote sensing
Deep learning

ABSTRACT

Drought is a popular scientific issue in global climate change research. Accurate monitoring of drought has important implications for the sustainable development of regional agriculture in the context of increasingly complex global climate change. Deep learning is a widely used technique in the field of artificial intelligence. However, ongoing on drought monitoring using deep learning is relatively scarce. In this paper, the various hazard factors in drought development were comprehensively considered based on satellite data including Moderate Resolution Imaging Spectroradiometer (MODIS) and tropical rainfall measuring mission (TRMM) as multi-source remote sensing data. By using the deep learning technique, a comprehensive drought monitoring model was constructed and tested in Henan Province of China as an example. The results showed that the comprehensive drought model has good applicability in the monitoring of meteorological drought and agricultural drought. There was a significant positive correlation between the drought indicators of the model output and the comprehensive meteorological drought index (CI) measured at the site scale. The consistency rate of the drought grade of the two models was 85.6% and 79.8% for the training set and the test set, respectively. The correlation coefficient between the drought index of the model and the standard precipitation evapotranspiration index (SPEI) was between 0.772 and 0.910 ($P < 0.01$), which indicated a strong level of significance. The correlation coefficient between the drought index of the model and the soil relative moisture at a 10 cm depth was greater than 0.550 ($P < 0.01$), and there was a good correlation between them. This study provides a new method for the comprehensive assessment of regional drought.

1. Introduction

Drought is one of the most serious meteorological disasters worldwide. As a common and frequently occurring disaster, it poses a serious threat to agricultural production, the ecological environment, and economic and social development (Dai, 2011). Therefore, research on drought monitoring techniques and assessment methods has important practical significance for the government's ability to respond to natural disasters. Current methods for monitoring drought include traditional meteorological monitoring methods (Dai et al., 2004) and remote sensing monitoring methods (Wang, 2004). Although the meteorological monitoring method is well established, the response of surface vegetation to drought has not been considered in the monitoring mechanism, therefore, such monitoring methods are limited in comprehensive drought monitoring. Remote sensing drought monitoring has the advantage of being macroscopic, rapid and providing continuous data in time and space. (Quiring and Ganesh, 2010).

Traditional remote sensing drought monitoring models mainly monitor single factors such as vegetation growth or soil moisture, which cannot fully reflect the information about drought. In recent years, many scholars have considered integrating multiple factors that are symptomatic of drought to build a comprehensive drought model (Yin et al., 2018; AghaKouchak et al., 2015). One of the methods of constructing such a model is traditional regression. Rhee et al. proposed the drought index, the scaled drought condition index (SDCI), of multi-source remote sensing data based on a linear combination of data (Rhee et al., 2010). Wang et al. used polynomial equations to fit the feature space of the normalized vegetation index (NDVI) and land surface temperature (LST) and established an enhanced temperature vegetation dryness index (ETVDI) to monitor regional drought (Wang et al., 2018). With the popularity of machine learning, some scholars have tried to use data mining methods to build drought models. Du et al. used the

* Corresponding author at: School of Geographical Sciences, Nanjing University of Information Science and Technology, NO. 219 Ningliu Road, Nanjing 210044, China.
E-mail address: rpshen@nuist.edu.cn (R. Shen).

https://doi.org/10.1016/j.jag.2019.03.006
Received 21 August 2018; Received in revised form 18 February 2019; Accepted 4 March 2019
Available online 08 March 2019
0303-2434/ © 2019 Elsevier B.V. All rights reserved.

method of classification and a regression tree to consider the combination of multiple drought factors such as vegetation, precipitation, and land surface, with terrain and land use data to construct a comprehensive drought index (SDI) (Du et al., 2013). Shen et al. used the method of random forests to consider the TRMM-Z index, the vegetation condition index (VCI), the temperature condition index (TCI) and other remote sensing drought indices to construct a comprehensive drought monitoring model (Shen et al., 2017). The development of these models provides new ideas for drought monitoring.

However, traditional regression methods primarily require multiple trials to determine weights to build a model, and these methods consider the drought-causing factors to be limited. Some common data mining methods also have limitations when handling large amounts of data. These problems may be solved with more research. Deep learning is a neural network-based machine learning method proposed by Hinton et al. in 2006 (Hinton and Salakhutdinov, 2006). It is one of the most popular areas in machine learning research. Deep learning can imitate the operation of the human brain to interpret data, and its performance surpasses other machine learning methods (Lecun et al., 2015). In the construction of comprehensive drought models, deep learning algorithms can extract more useful features from a large number of drought factors, which is beyond the reach of other traditional algorithms. However, there are few studies on drought monitoring using deep learning. Therefore, this study used deep learning methods to construct models by considering a number of various hazard factors and explored the use of multiple remote sensing data sources for regional remote sensing comprehensive drought monitoring methods.

2. Materials

2.1. Study area

Henan Province is located in the hinterland of the North China Plain between 110°22′~116°38′E and 31°23′~36°22′N (Fig. 1). This area has a transitional climate with both humid subtropical climate and semi-humid monsoon climate with warm temperate zones. The average annual precipitation is 500–1000 mm, and the precipitation season is unevenly distributed. 50% of the annual precipitation is concentrated in the summer. The terrain is high in the west and low in the east, with mountains in the north, west, and south, plains in the east, basins in the southwest, and four valleys across the Yellow River, the Haihe River, the Huaihe River, and the Yangtze River (Shen et al., 2017). The main crops in Henan Province are winter wheat and summer maize, and the southern crop is rice.

2.2. Remotely sensed data

The Moderate Resolution Imaging Spectroradiometer (MODIS) vegetation index product (MOD13A3), surface temperature product (MOD11A2), land use product (MCD12Q1), and Tropical Rainfall Measuring Mission (TRMM) product were used as the main remote sensing data sources from 2001 to 2013. MOD13A3 is a monthly synthetic surface vegetation index product, and MOD11A2 is a land surface temperature product that is synthesized every 8 days. Both of the products have a spatial resolution of 1 km. MCD12Q1 is a yearlong-synthesized land cover type product with a spatial resolution of 500 m. TRMM 3B43 is a 0.25° × 0.25° monthly average grid precipitation data set (mm/h). In this research, TRMM data and MODIS data were downloaded from the website (https://search.earthdata.nasa.gov/). In addition, the study also used the Shuttle Radar Topography Mission Digital Elevation Model (SRTM-DEM)data obtained by the International Centre for Tropical Agriculture (CIAT) using the new interpolation algorithm from the website (https://ciat.cgiar.org/).

2.3. Meteorological and soil data

The meteorological data utilized in this study includes monthly average temperature, precipitation data, and soil relative moisture from 15 major meteorological stations and 9 agricultural meteorological stations in Henan Province (Fig. 1), which was downloaded from the website (http://data.cma.cn/).The data has been quality controlled. In addition, in order to calculate the soil available water capacity (AWC), a data set of Chinese soil texture distribution published by Beijing Normal University was adopted, with a spatial resolution of 1 km (http://globalchange.bnu.edu.cn/).

3. Methods

3.1. Data processing

3.1.1. Remote sensing data

TRMM 3B43 data was converted to monthly precipitation and resampled to 1 km resolution by bilinear interpolation (Du, et al., 2013). NDVI and enhanced vegetation index (EVI) were extracted from the MOD13A3 data. Quality control documents were used to eliminate invalid values from the images and the percentage of missing pixels in all NDVI and EVI data used in the study was 0.044% and 0.024% respectively. The multi-year averages from the same month in other years were used to fill where the data were removed. For the MOD11A2 8-day land surface temperature data, all monthly data were weighted and added to obtain the surface temperature monthly value. The weight was taken as the proportion of the number of days per scene image in the month. In this study, the use of MOD11A2 8-day data avoided a large number of missing values in the daily data, and greatly reduced the amount of data required for model.

3.1.2. Meteorological data

In this study, the comprehensive meteorological drought index (CI) and the standardized precipitation evapotranspiration index (SPEI) were calculated based on the monthly precipitation and mean air temperature at each meteorological station from 2001 to 2013. SPEI uses the degree of difference between precipitation and evapotranspiration to determine the deviation from the average state to characterize the drought in a certain region. This study used the method of Vicente-serrano to perform the calculation (Vicente-serrano et al., 2010). The CI is based on the standardized precipitation index, wetness index, and recent precipitation. It is superior to the drought index that is calculated by only using precipitation. Details of the calculation method can be found in the description of He's literature (He et al., 2015).

3.1.3. Soil data

Soil relative moisture is an important measure of the impact of drought on agriculture. The soil relative moisture at a depth of 10 cm from 2001 to 2013 was obtained from nine agricultural meteorological stations in the study area. The monthly relative moisture data of the 10 cm deep soil layers at each site was calculated after quality control. The main principle of quality control was that soil moisture observers failed to work normally below 0 °C.Therefore, according to the soil temperature observation at a depth of 10 cm, the corresponding data was filtered. If the soil temperature at a depth of 10 cm was less than 0 °C, the corresponding soil relative moisture data was removed.

Soil Available Water Capacity (AWC) refers to the amount of water stored in the soil that can be used by plants. In this study, the empirical linear fitting model using soil texture calculation AWC proposed by Petersen (Gupta and Larson, 1979) was used to estimate the effective water holding capacity in a 10 cm deep soil layer.

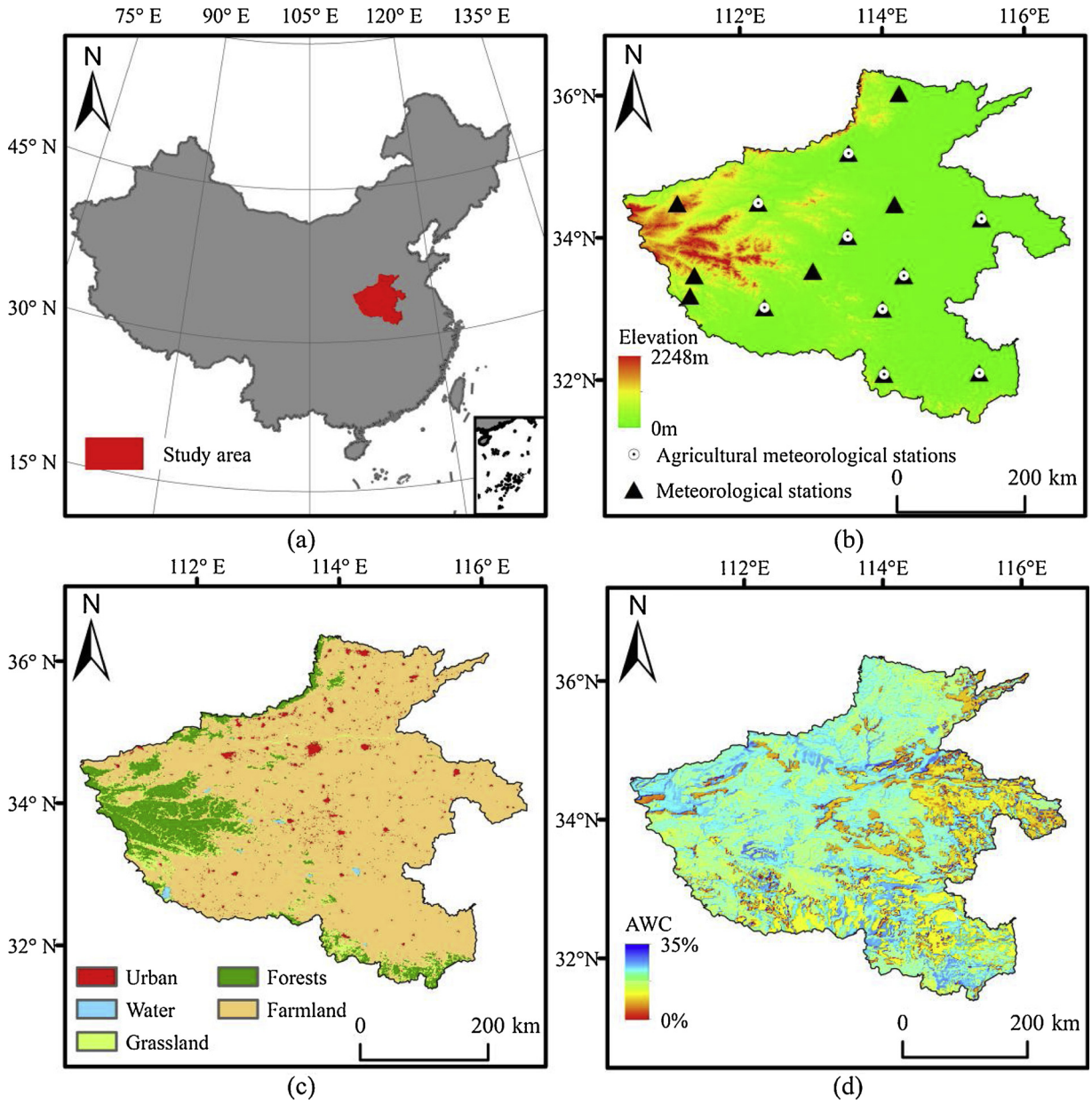

Fig. 1. Maps of the study area conditions. (a) Location map of the study area. (b) Elevation map of the study area and location map of the meteorological and agricultural stations. (c) Land cover map of the study area in 2013. (d) AWC map of the study area.

3.2. Drought monitoring model construction

3.2.1. Principles of model construction

The theoretical basis for the composite drought monitoring model is that drought is comprehensively determined by a variety of drought factors, that are not only related to precipitation, soil water stress, and vegetation growth status but are also related to factors such as available water capacity, land cover types, and landform types (Du et al., 2013). The Vegetation Condition Index (VCI) reflects the growth status of the vegetation, and the Temperature Condition Index (TCI) reflects the influence of surface temperature on vegetation growth (Kogan, 1997). The Vegetation Supply Water Index (VSWI) reflects the situation of the water stress of plants. TRMM-Z index and Percentage of Precipitation Anomaly (Pa) reflect the information of the meteorological precipitation anomaly. The Available Water Capacity (AWC) reflects the influence of different soils on drought. Land cover types (LC) and elevation also have important influences on regional drought. Each factor reflects the drought in different aspects, but the manner in which they are coupled to drought is still unclear. At present, the CI not only reflects the anomaly of monthly and seasonal scales of precipitation but also has a certain ability to monitor the short-term scale of water deficit. Therefore, this study used the method of deep learning, which uses CI as a dependent variable and other remote sensing factors as independent variables, to construct a drought monitoring model that is driven by multi-source remote sensing data; through this work, a better drought monitoring method is explored.

3.2.1.1. Vegetation condition index. When vegetation is under drought stress, the NDVI value will be reduced accordingly. However, a single NDVI image only reflects the relative health of the vegetation growth at a specific time. Therefore, in the construction of the model in this the study, the VCI that can reflect the vegetation growth on the time series was chosen (Liu and Kogan, 1996). The VCI for each month from 2001 to 2013 was calculated by the following formula:

$$VCI_i = \frac{NDVI_i - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}} \times 100 \quad (1)$$

where VCI_i is the vegetation state index of a certain month, $NDVI_i$ is the NDVI value of a certain month, and $NDVI_{\min}$ and $NDVI_{\max}$ are the minimum and maximum values of NDVI in the corresponding month in the research year. Lower VCI value indicates worse vegetation growth status. The index reduces the noise impact of seasonal changes via a ratio.

3.2.1.2. Temperature condition index. The occurrence and development of drought is closely related to LST. High LST during the vegetation growing season indicates unfavourable or drought conditions, while low LST indicates mostly healthy conditions (Singh et al., 2003). However, LST is affected by many factors such as the atmosphere and the environment. Therefore, the use of LST alone cannot monitor drought completely. The TCI is defined based on the principle that the surface temperature of the canopy or soil increases with the increase of water stress under high temperature conditions that cause vegetation or water shortages in the soil. The TCI focuses on the threat of high temperature to the growth of vegetation (Kogan, 1995). This study calculated the TCI for each month between 2001 and 2013 and used it as one of the model's input variables. The formula is as follows:

$$TCI_i = \frac{LST_{\max} - LST_i}{LST_{\max} - LST\min} \times 100\% \quad (2)$$

where TCI_i is the temperature condition index of a certain month, LST_i is the LST value of a month, and $TCI_{\max}$ and $TCI_{\min}$ are the maximum and minimum values of LST for the corresponding month in the study year. Smaller TCI means more severe drought.

3.2.1.3. Vegetation supply water index. The VSWI combines the information of vegetation status and surface temperature. The physical meaning is clear, and the form is simple. The physical meaning is that when the crop water supply is normal, the remote sensing vegetation index and crop canopy temperature remain within a certain range during a certain growth period. Internally, if a crop is affected by a drought or water shortage, the vegetation index decreases. During this time, the crop does not have enough water to evaporate on the surface of the leaf and it is forced to close some of the stomata, causing the temperature of the crop canopy to rise. Li et al. proposed that the construction of VSWI using EVI is effective in Henan Province (Li et al., 2016). This study calculated the VSWI for each month from 2001 to 2013.The formula is as follows:

$$VSWIi = \frac{EVI_i}{LSTi} \quad (3)$$

where $VSWI_i$ is the Vegetation Supply Water Index of a certain month, EVI_i is the EVI value of a month, and LST_i is the LST value of a month. Smaller VSWI indicates more severe drought.

3.2.1.4. Percentage of precipitation anomaly and TRMM-Z index. Precipitation anomaly is one of the indicators used to characterize the drought caused by precipitation anomalies in a certain period. It is one of the indicators used in daily meteorological operations to assess the amount of water relative to the normal value. It can directly reflect the occurrence of droughts in months, seasons, and years (Ren, 2013). In this paper, the Pa was used to standardize the precipitation anomaly. The monthly Pa is calculated as follows:

$$Pai = \frac{Pi - \overline{P}}{\overline{P}} \times 100\% \quad (4)$$

where Pa_i is the percentage of precipitation anomaly for a month, P_i is the precipitation in a month, and $\overline{P}$ is the average annual precipitation in a month.

In general, the precipitation in a certain period of time does not obey a normal distribution. It is better to use the theoretical curve of Pearson III distribution to fit the precipitation for a period of time (Vicente-Serrano, 2006). By normalizing precipitation value, the Pearson-III distribution can be converted into a standard normal distribution with Z as the variable (Kite, 1988). The formula is as follows:

$$Zi = \frac{6}{Cs}\left(\frac{Cs}{2}\varphi i + 1\right)^{\frac{1}{3}} - \frac{6}{Cs} + \frac{Cs}{6} \quad (5)$$

where Z_i is Z-index for a month, C_s is coefficient of skew, and φi is standardized precipitation value for a month. Cs and φi can be calculated from the precipitation sequence. The formulas are as follow:

$$C\,s = \frac{\sum_{i=1}^{n}(Pi - \overline{P})^3}{ne^3}\varphi i = \frac{Pi - \overline{P}}{e} \quad (6)$$

$$e = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Pi - \overline{P})^2\,\overline{P}} = \frac{1}{n}\sum_{i=1}^{n} Pi \quad (7)$$

where P_i is the precipitation in a month. Z-index is used to characterize the drought and flood levels of single-station meteorological data. To obtain the precipitation anomaly information in study area, we calculated the TRMM-Z index for each pixel using the TRMM precipitation data.

3.2.1.5. Other factors. The AWC measures the ability of a well-drained soil to supply water to plants. As an input parameter for various meteorological and remote sensing drought models, it has an important significance for the extraction of drought information. The occurrence of drought has differing impacts at the surface depending on the land cover type. In particular, there are great differences in the impact of drought for different land cover categories (such as agricultural and forest land). Therefore, referring to the method of Ran (Ran et al., 2010), the International Geosphere-Biosphere Programme (IGBP) classification scheme of MCD12Q1 data in Henan Province was reclassified into six categories which included farmland, woodland, grassland, water bodies, urban construction land and bare land. Different regional elevations (such as mountains and plains) have different drought developments. Therefore, elevation factors should also be considered when constructing drought models.

3.2.2. Deep learning

The concept of deep learning stems from the study of artificial neural networks. A multilayer sensor with multiple hidden layers is a deep learning structure. Deep learning can combine low-level features and then build up more comprehensive high-level attribute features in a layer-by-layer manner. Traditional neural networks usually have only two or three layers of neural networks. The parameters and computation nodes are limited. It has limited ability to learn and express complex functions. The deep learning model has five to ten layers or even more neural networks, and it results in a more effective training mechanism. The hierarchical structure of the deep learning model is only connected between adjacent layer neurons, and the same layer and cross-layer neurons are not connected to each other, which is similar to the structure of the human brain and can imitate the brain to represent information efficiently and accurately.

In this study, a deep feed forward neural network (DFNN) in the deep learning model was used as a research model. In regression tasks, it can extract high-level features among a large number of variables for high prediction accuracy (Zhang et al., 2017). This study used the deep learning framework in H2O, which is based on a DFNN that is trained with a gradient descent using error backpropagation. The H2O R package 3.0 edition can be found at the following link: https://www.h2o.ai. The specific training process for deep learning includes: (1) feature learning that uses a bottom-up, unsupervised approach. First, the first layer is trained with no calibration data, and the parameters of the first layer are learned. Because of the model capacity and sparse constraints, the resulting model can learn the data structure itself. The

model has more features than the input. After learning to obtain the n-1th layer, the output of the n-1 layer is taken as the input of the nth layer, and the nth layer is trained, thereby obtaining the parameters for each layer. This step is the most different from traditional neural networks. (2) The model uses top-down supervision learning. By training with tagged data, errors propagate backward from the top to the bottom, the network is fine-tuned, and the weight between layers is adjusted.

3.2.3. Process of model construction

The monitoring period for this study was selected from February to November for each year 2001-2012. April to October is the growing season for crops, and VCI and TCI are suitable for monitoring crops in the growing season. In order to study the interaction mechanism between precipitation, vegetation, and soil, and to obtain a quantitative relationship between this mechanism and comprehensive drought, a drought monitoring model with CI = f (VCI, TCI, VSWI, Pa, TRMM-Z, AWC, LC, DEM) was constructed based on deep learning. The monthly CI was used as the dependent variable, which indicated the comprehensive drought condition in a single station. VCI, TCI, VSWI, Pa, TRMM-Z index, LC, DEM, and AWC were used as the independent variables. VCI, TCI and VSWI represented vegetation-temperature factors, and Pa and TRMMZ represented precipitation factors. AWC, LC and DEM represented soil factors, land cover factors and topographic factors respectively. This study used data from 2001 to 2010 as a training set and data from 2011 to 2012 as a test set. The model building procedure, as shown in Fig. 2, included the following steps: (1) calculate the parameters needed by the model and convert them into data sets of the same format as the input parameters. (2) Adjust the training parameters of the model, including setting the number of hidden layers, the number of hidden units in each hidden layer (neurons) and the number of iterations on the training sample, etc. (3) Use the DFNN training model to achieve model optimization by adjusting the hidden layers, neurons, and number of iterations, using cross-validation between the test set and the training set.

Table 1
A selection of results of tuning the key parameters of the model.

Layers	Neurons	Epochs	Train-R^2	Train-RMSE	Test-R^2	Test-RMSE
7	500	1000	0.751	0.501	0.702	0.756
7	500	2000	0.771	0.450	0.721	0.678
7	500	3000	0.799	0.432	0.756	0.621
8	500	1000	0.855	0.382	0.778	0.482
8	500	2000	0.890	0.295	0.825	0.357
8	500	3000	0.882	0.282	0.808	0.421
9	500	1000	0.865	0.366	0.772	0.440
9	500	2000	0.899	0.285	0.802	0.391
9	500	3000	0.872	0.296	0.797	0.401
…	…	…	…	…	…	…

4. Results and discussion

4.1. Model calibration and validation

By tuning the number of hidden layers, neurons and epochs, the coefficient of determination (R^2) and root mean square error (RMSE) of the model were obtained, and a selection of the results are shown in Table 1. When the highest R^2 value and the lowest RMSE value of the training set occurred, the R^2 and RMSE of testing set were not optimal. This result suggested that blindly pursuing the minimization of training set errors could lead to a decrease in the ability to predict unknown data (test set). This trend was generally exhibited in models with too many parameters. Therefore, to avoid over-fitting and to improve the generalization ability of the model, the focus was on the evaluation parameters of the training set and testing set simultaneously. In this study, we ultimately chose the following as the initial input parameters of the model: hidden layers = 8, neurons = 500 and epochs = 2000.

The optimal version of the model was validated using both the training set and the testing set. As shown in Table 1, the R^2 between the estimated drought index (DI) value of the training set and the measured CI value is 0.890, and the R^2 between the estimated DI value of the testing set and the measured CI value is 0.825. In addition, the RMSE of the training set and the test set were 0.295 and 0.357, respectively. This result indicates that the simulation accuracy of the model is higher and the prediction ability for new data is better than before.

With reference to the criteria for the classification of drought grades

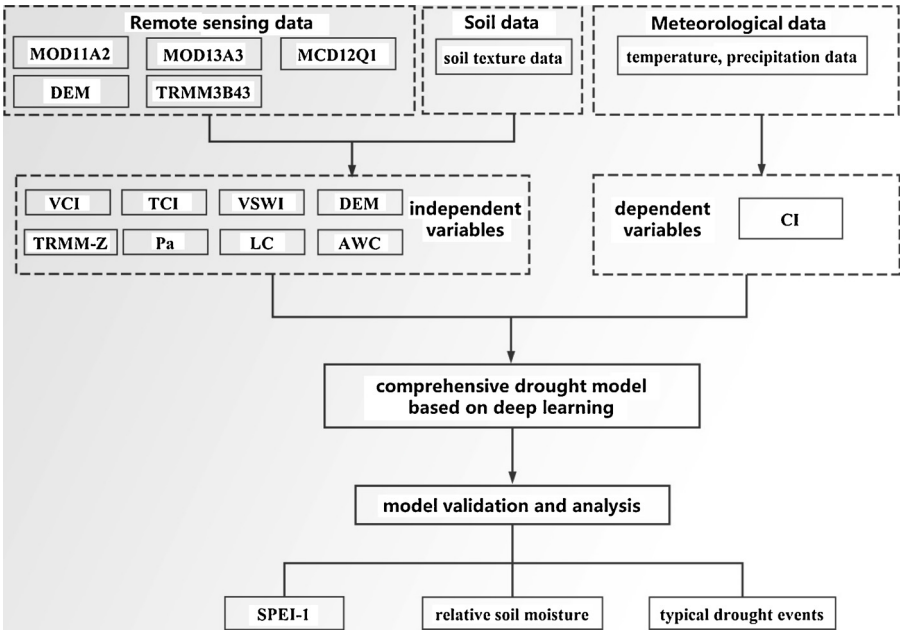

Fig. 2. Flowchart of drought monitoring model construction.

Table 2
The criteria for the classification of drought grades in CI.

Grade	CI value
normal	$-0.6 < CI$
mild	$-1.2 < CI \leq -0.6$
moderate	$-1.8 < CI \leq -1.2$
severe	$-2.4 < CI \leq -1.8$
extreme	$CI \leq -2.4$

Table 3
Accuracy analysis of model prediction in the training and testing sets.

	R^2	RMSE	Consistency rate	Vacancy rate	Miss rate
Training set (2001-2010)	0.890	0.295	85.6%	7.1%	7.4%
Test set (2011-2012)	0.825	0.357	79.8%.	5.1%	8.9%

in CI (Zou et al., 2010), the estimated values were classified into extreme drought, severe drought, moderate drought, mild drought and normal (Table 2). In this paper, the consistency rate of the drought grades between the measured CI and the estimated value of the model was calculated. As shown in Table 3, the consistency rate between the measured CI value and the training set estimated value is 85.6% and that between the measured CI value and the testing set estimated value is 79.8%. Furthermore, the vacancy and the miss rate of the drought grades between the measured CI and the estimated value were also counted, and these results are presented in Table 3. Both the vacancy rate and the miss rate are less than 10%, which proves that the model is sensitive to drought events and has good monitoring of drought events.

4.2. Correlation analysis based on meteorological drought

To verify the ability of the model to monitor meteorological drought information, this study used the standard precipitation evapotranspiration index (SPEI) that was calculated from meteorological station observation data to verify the comprehensive drought index estimated by the model. SPEI includes two major factors, precipitation and temperature, which not only can sensitively reflect the changes in evaporation caused by temperature fluctuations but also has the advantages of simple calculation and multiple time scales. It has been widely used in drought monitoring (Vicente-serrano et al., 2010). In areas where China's average annual precipitation is greater than 200 mm, SPEI over various time scales have a better applicability (Wang et al., 2015). In this study, we calculated the SPEI on a one-month time scale (SPEI-1) for 15 meteorological stations in Henan Province from February to November in 2001-2012. As shown in Fig. 3, a correlation analysis is performed between the comprehensive drought index estimated by the model and SPEI-1. The results showed that the comprehensive drought index obtained by the model has a strong correlation with SPEI-1. In addition to their correlation coefficient of slightly less than 0.8 in June, correlation coefficients were higher than 0.8 in the other study months, and all of the months passed the $P < 0.01$ significance test. As the main food crops in Henan, wheat and corn are the main culprits in meteorological drought. The growth period of wheat primarily occurs from March to May, and the growth period of corn primarily occurs from July to September. The comprehensive drought index obtained by the model had a significant correlation with the meteorological drought index during the growth period of these food crops, which indicated that the model has a good potential in meteorological drought monitoring.

4.3. Correlation analysis based on relative soil moisture

To verify the applicability of the constructed model in monitoring agricultural drought, nine homogeneously distributed agricultural meteorological sites were selected based on the location of the agricultural meteorological sites in Henan and the type of crop planting. By accounting for the relative soil moisture of these sites from March to November in 2001–2012, we analysed the correlation between the monthly soil relative moisture at a depth of 10 cm at each site and the monitoring value of the drought monitoring model using deep learning. The scatter plots of the monitoring results and the measured relative soil moisture are shown in Fig. 4. The results showed that the correlation coefficients of all sites were all above 0.5, and they all passed the significance test of $P < 0.01$. There was a good correlation between the monitored values and the relative soil humidity. The comprehensive drought index obtained by the model can reflect the change in soil moisture information to some extent. Because soil moisture is the decisive factor of agricultural drought, the comprehensive drought index obtained by the model can be applied to regional agricultural drought monitoring.

The monitoring effect of agricultural drought in this paper was slightly lower than that of meteorological drought. It is mainly because the monitoring result is 1 km resolution, and at this scale, the spatial heterogeneity of relative soil moisture is usually greater than that of precipitation, which results in the site relative soil moisture not fully representing the soil moisture at 1 km pixel scale, thus weakening the correlation between relative soil moisture and model monitoring results. In addition, the influence of human factors such as irrigation under drought conditions can cause relative soil moisture to fluctuate, thus reducing the reliability of model verification using the relative soil moisture.

4.4. Analysis of model monitoring results

According to the National Meteorological Drought and Flood Distribution Data Released by the National Climate Center of the China Meteorological Administration (http://www.cmdp.ncc-cma.net/) and the Henan Province Agricultural Meteorological Monthly Report (http://www.henan.weather.com.cn/), there was a significant drought from August to November in 2013 in Henan. In August, the monthly average temperature across the province was 1–4 °C higher than in the same period in a normal year; most of the monthly precipitation was almost 20% to 90% less than in a normal year. In September, temperature was high in most of the regions, which was unfavourable for the upcoming winter wheat planting. In October, the monthly precipitation in most other areas was 20% to 90% less, and Henan had drought conditions in the western, central, and northern regions, which caused adverse effects on the seedling growth of winter wheat and rapeseed. In November, except for below average precipitation in the southwest, the overall precipitation in other areas was relatively high. The drought in most areas was eased, and the drought period was basically over.

In this study, the drought monitoring model based on deep learning was used to monitor and classify drought from August to November in 2013, and the results are shown in Fig. 5(e)–(h). The results showed that serious drought conditions occurred in the northern and western parts of Henan Province in August, and severe drought and moderate drought occurred in most of the areas. Extreme drought occurred in some areas, and there was essentially no drought in the central and southern regions. The main reason was that the temperature in the province was high in August, and the precipitation in the northern region was 20% to 90% less than in previous years, while precipitation in the southern region was sufficient. In September, the drought in the north intensified, resulting in a large area of severe drought and extreme drought, and the drought spread to the central and eastern regions. At the same time, there was no drought in southern Henan,

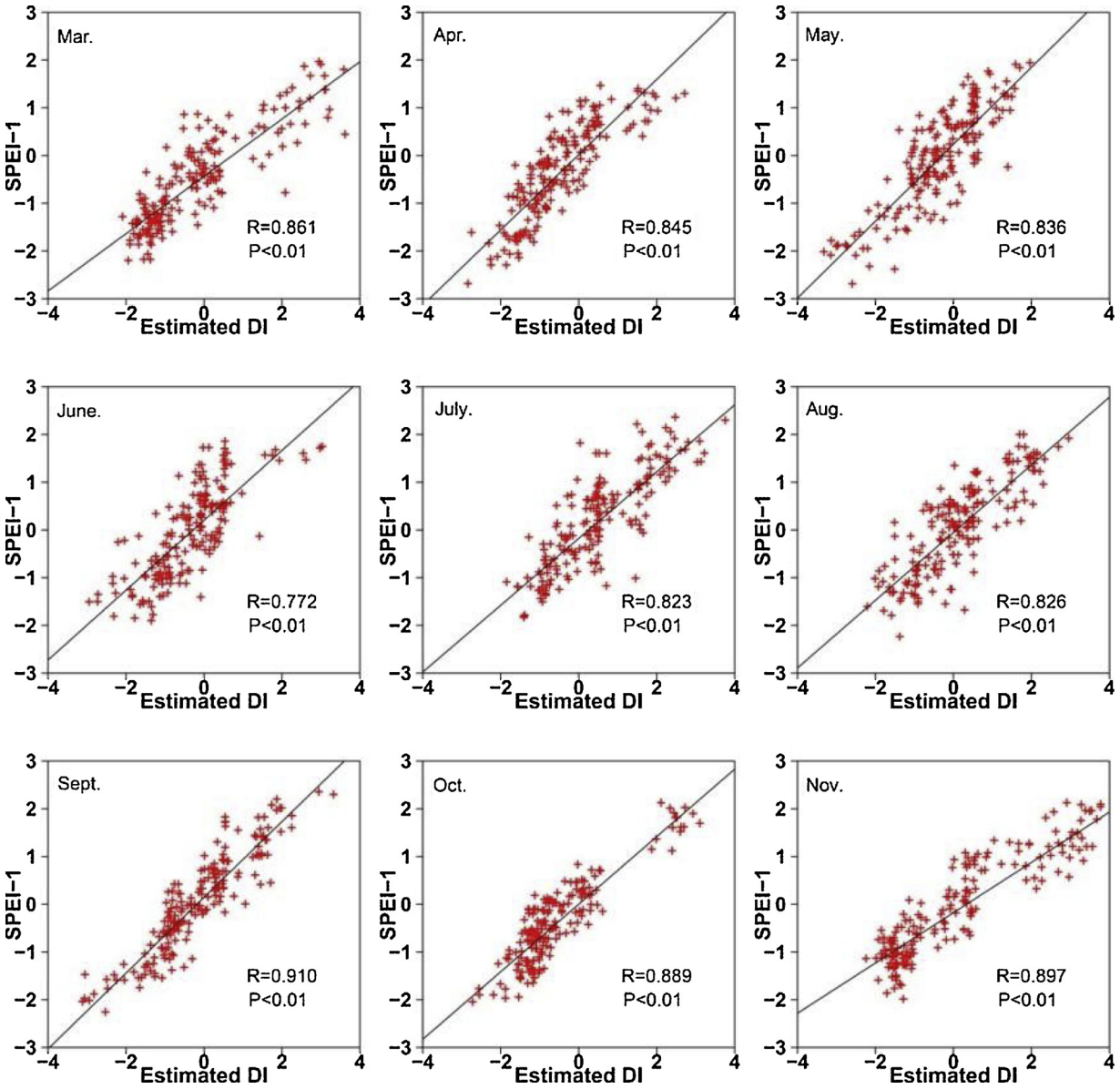

Fig. 3. Scatter diagrams of the drought index estimated by model and the SPEI for the one-month time scale.

mainly because the precipitation in most parts of southern Henan was more than 20% to 140% higher than in the same period in normal years. In October, there was a drought in most parts of the province, which was mainly related to high temperatures and low precipitation in most areas. In the south, northwest and northeast regions, there was a moderate or more severe drought. However, overall, drought had a tendency to decrease in severity. In November, except for a mild drought in the southwestern region, the drought in other areas had basically stopped. The main reason for this is that there were more precipitation processes in the province this month, which had a positive impact on the mitigation and elimination of drought conditions. This was beneficial to the winter wheat tillering and the formation of pre-winter strong seedlings. At this point, the drought process was also concluded. Combined with the above analysis process, the remote sensing drought monitoring model based on deep learning that was established in this study can be used to monitor the temporal and spatial evolution of the drought process in Henan Province. The results were basically consistent with the actual drought situation, which indicates that the model has a relatively strong monitoring capability.

Currently, in the meteorological service system, CI is calculated from data observed by a single station, and it lacks spatial continuity. Usually, meteorological monitoring methods deal with the problem of spatial discontinuity of observation data by interpolation. In this paper, inverse distance weighting interpolation (IDW) was used to convert the CI value calculated by station observation data from station scale to regional scale (Fig. 5). The results showed that the results of station interpolation generally reflected the process of drought development. However, due to the limited number of meteorological stations, the results could not effectively reflect the details of the drought distribution and lack credibility in places where there was no station. For example, In September to October, the interpolated results aggravated the drought in the western forest region. It is because the interpolation uses a simple mathematical method, without considering the influence of other factors such as vegetation type on the drought. In addition, the distribution of meteorological stations in forest areas is extremely small, leading to inaccurate interpolation results. In November, There was still mild drought in the southwest of Henan Province because of below average precipitation. The interpolation results failed to monitor the drought because of only one station in this area, while the model results reflect the drought in this area. The monitoring model

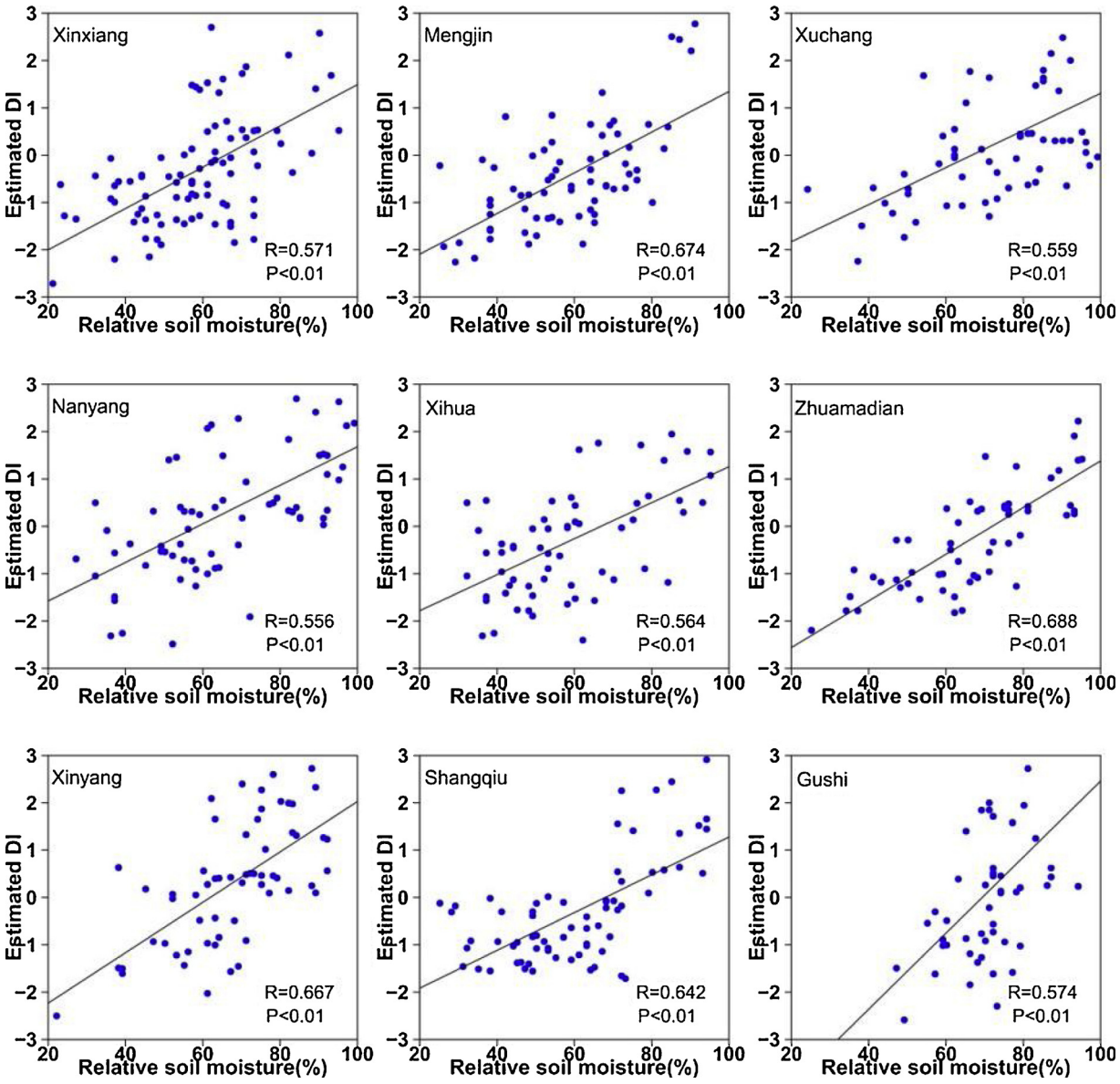

Fig. 4. Scatter diagrams of the drought index estimated by the model and the relative moisture of 10 cm soil depth.

established in this paper used high-resolution remote sensing data as input data, so the monitoring results were more detailed than the interpolation results. Excellent local detail features made this model better than traditional meteorological monitoring methods.

Through the analysis of the monitoring results of the model and comparison with traditional meteorological monitoring methods, the model constructed in this paper has the following advantages: Firstly, The model constructed in this paper considered various drought stress factors, including soil water stress, vegetation growth status and meteorological precipitation loss. Therefore, the model constructed in this paper has the ability to monitor multiple types of drought. Secondly, Traditional meteorological interpolation methods require the construction of a large number of meteorological stations in order to obtain a higher spatial precision result. The model constructed in this paper used free remote sensing data and a small amount of meteorological station data, which not only obtained high-resolution monitoring results, but also cost less. Moreover, When dealing with big data, the deep learning method is superior to the traditional machine learning method in the ability of feature discovery and combination (Zhang et al., 2017), which is the reason why the model constructed in this papercould accurately mine the comprehensive drought information in multi-source remote sensing data.

4.5. Influences of different impact factors on simulation

In this paper, 8 influence factors were used as independent variables for model input. These factors were considered equally important at the input layer of the model, but they had different contributions to the final simulation results. To study the influences of different impact factors on simulation, we obtained the contribution rate of different factors to the simulation from the model calculation results (Table4). The results showed that the contribution rates of TRMM-Z and Pa were higher than other factors, which means that precipitation was the most important factor affecting the drought in this model. The contribution rates of VCI, TCI and VSWI were closed to 0.9, which means vegetation factors and surface temperature factors also played an important role in the formation and development of drought in this model. It is mainly because the transpiration of healthy vegetation has a cooling effect on the land surface, which can reduce the loss of soil moisture. When the vegetation grows poorly or the surface temperature is high, the loss of soil moisture will increase, resulting in regional drought. AWC also had a contribution rate greater than 0.8. AWC reflects the soil moisture

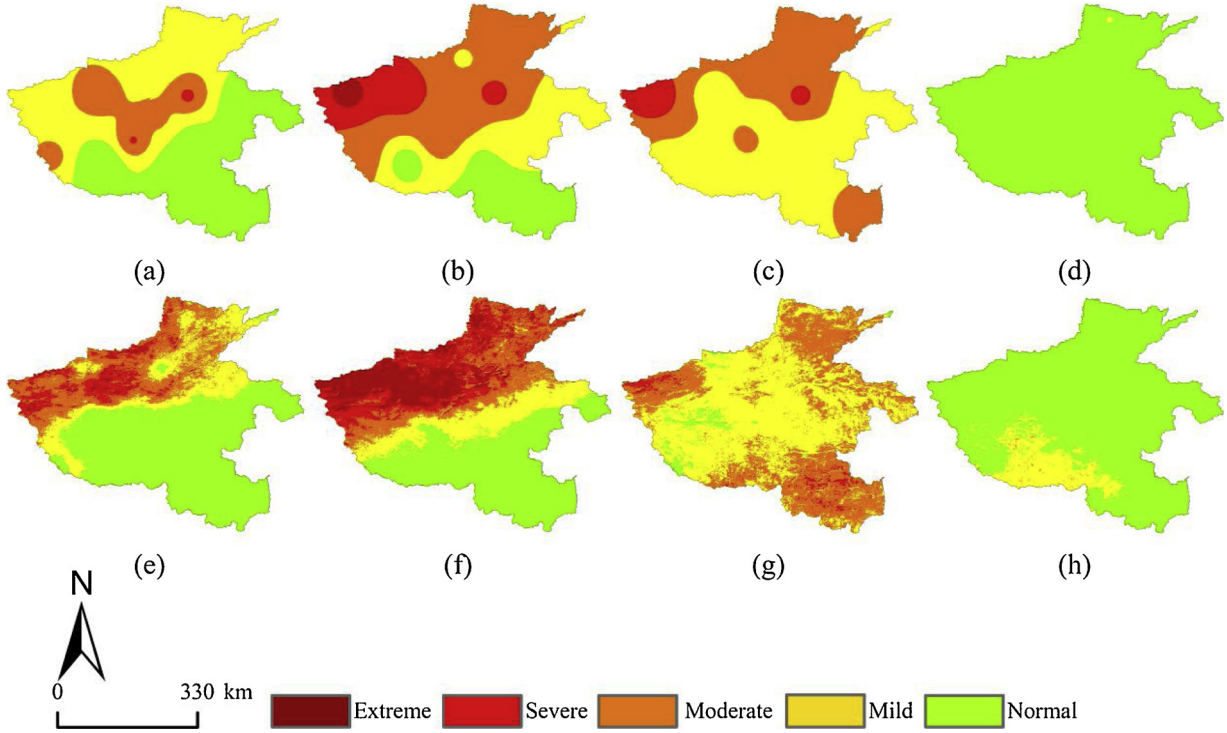

Fig. 5. Drought maps of Henan province from August to November in 2013. (a)˜ (d) Drought maps of Henan province from August to November in 2013 by using IDW. (e)–(h) Drought maps of Henan province from August to November in 2013 by using deep learning model.

Table 4
the contribution rate of different factors to the simulation.

Drought factors	Contribution rate
TRMM-Z	0.957
Pa	0.935
VCI	0.899
TCI	0.894
VSWI	0.889
AWC	0.823
LC	0.646
DEM	0.569

content that can be absorbed by plant roots. High AWC indicates that the soil has strong drought resistance.

The contribution rates of LC and DEM to the model were lower than other factors. We calculated the proportion of land cover types in the study area, and the results showed that the land cover in the study area was relatively simple. As of 2013, farmland and forest accounted for 83.6% and 11.7% of the total area respectively, while other land cover types accounted for less than 5%. These land cover types changed less in the past 12 years, with only about 3.6% of the total area of the grassland becoming farmland. Simple land cover types reduced the impact of LC on the model simulation in this study. In general, forests have much higher water conservation capacity than farmland, which leads to a lower frequency and extent of drought in forest-covered areas. For example, In September to October, drought developed slowly and more easily disappeared in western forest areas than in farmland. Therefore, the implementation of appropriate returning farmland to forests in the region can reduce the occurrence of drought. In this model, the contribution rate of DEM was the lowest, indicating that in this study area, drought was more directly affected by other factors than by elevation. Although elevation had the lowest contribution rate, it still had an indirect effect on drought. For example, precipitation in the mountains generally increases with elevation and reaches a maximum at a certain elevation, which is conducive to reducing the possibility of drought in mountainous areas. In addition, the undulating terrain leads to the large slope increases the likelihood of soil moisture loss, making the drought more likely to occur. Therefore, reducing forest resource development in hills can reduce the occurrence of regional drought events.

4.6. The limitation of the model for drought monitoring

The model constructed in the paper still had some limitations that needed to improve in future work. The following are main limitations and future developments:

(1) In this paper, we mainly used the deep learning method to mine drought information from remote sensing big data to explore a new method for regional drought monitoring using multi-source remote sensing data. In order to prove the model constructed in the paper is superior to other drought monitoring tools, a further comparison between this model and other models such as support vector machines, random forests, or enhanced decision trees is needed in future work.
(2) In the construction of the model, considering the model ability to rebuild and monitor drought for a long time, we chose TRMM which has the longest time series in current satellite precipitation products. The principle of the precipitation index construction in this paper was to regard the pixel of remote sensing images as a meteorological station. Similarly, for other satellite precipitation products such as GPM and CMORPH, the precipitation index can be constructed by the same method. However, due to the different precision of TRMM, GPM and CMORPH in different regions of China (Lu et al., 2018), the model simulation results can be affected to some extent. In future research, we plan to use TRMM, CMORPH and GPM data to produce a set of long-term multi-source remote sensing precipitation data at a national scale for national scale simulations, the most critical technology of which is the downscaling of precipitation data, especially the precipitation downscaling under complex terrain such as mountains.

(3) The deep learning model has the ability to high-level features from a large number of low-level features. For this model, the more input factors indicate that it can mine more drought information. Therefore, more impact factors based on remote sensing data such as soil moisture data, evapotranspiration data and surface albedo data should be studied in future research. Furthermore, the deep learning model framework constructed in this paper can be implemented based on R or Python and owns good portability, which provides the basis for our future research to transfer the developed technique to the operational applications.

5. Conclusions

The drought process involves many factors such as atmosphere, soil, and vegetation. Only by comprehensively considering the model's internal coupling process and the drought stress factors, including soil water stress, vegetation growth status and meteorological precipitation loss, can we accurately describe and evaluate the true drought conditions. In this study, multi-source spatial data and deep learning methods were used to comprehensively analyze the effects of a large number of drought factors on drought, and an integrated drought monitoring model based on deep learning was constructed. The drought from March to November in Henan Province from 2001 to 2012 was analysed. The following is a summary of the results.

(1) The comprehensive drought monitoring model can quantitatively monitor the regional drought. By establishing cross-validation between the training set and the test set during the model training process, the generalization ability of the model was improved. The consistency rate of the drought level obtained from the model training set and the measured CI drought level reached 85.6%, and the consistency rate of the drought level obtained from the model test set and the measured CI drought level reached 79.8%.
(2) The comprehensive drought model can not only reflect the degree of meteorological drought but also monitor agricultural drought. From March to November, the correlation coefficient between the comprehensive drought index and the standardized precipitation evapotranspiration index (SPEI) obtained by the model was above 0.8 ($P < 0.01$), except for that in June, and reached a significant correlation level. There was also a certain correlation between the model and the soil moisture at a 10 cm depth at various agricultural meteorological sites, and the correlation coefficient was higher than 0.5 ($P < 0.01$). The model had good applicability to comprehensive drought monitoring.
(3) Monitoring the drought events in Henan Province from August to November 2013 based on the model was performed in this study. The monitoring results were consistent with the actual drought conditions, and they reflect the development and spatial evolution of drought conditions.

Funding

This research was supported by the Key Project of the National Natural Science Foundation of China (Grant No.91437220) and the National Key R&D Program of China (2018YFC1506602).

References

AghaKouchak, A., Farahmand, A., Melton, F.S., Teixeira, J., Anderson, M.C., Wardlow, B.D., Hain, C.R., 2015. Remote sensing of drought: progress, challenges and opportunities. Rev. Geophys. 53, 452–480.
Dai, A., 2011. Erratum: drought under global warming: a review. Wiley Interdiscip. Rev. Clim. Change 2 (1), 45–65.
Dai, A., Trenberth, K.E., Qian, T., 2004. A global dataset of Palmer Drought Severity Index for 1870 - 2002: relationship with soil moisture and effects of surface warming. J. Hydrometeorol. 5 (6), 1117–1130.
Du, L.T., Tian, Q.J., Yu, T., et al., 2013. A comprehensive drought monitoring method integrating MODIS and TRMM data. Int. J. Appl. Earth Obs. Geoinf. 23 (8), 245–253.
Gupta, S.C., Larson, W.E., 1979. Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density. Water Resour. Res. 15 (6), 1633–1635.
He, J., Yang, X.H., Li, J.Q., et al., 2015. Spatiotemporal variation of meteorological droughts based on the daily comprehensive drought index in the Haihe River basin. China. Natural Hazards 75 (2), 199–217.
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507.
Kite, G.W., 1988. Frequency and Risk Analyses in Hydrology. Water Resources Publications, Littleton, Colorado.
Kogan, F.N., 1995. Application of vegetation index and brightness temperature for drought detection. Adv. Space Res. 15 (11), 91–100.
Kogan, F.N., 1997. Global drought watch from space. Bull. Amer. Meteor. Soc. 78, 621–636.
Lecun, Y., Bengio, Y., Hinton, G.E., 2015. Deep learning. Nature 521 (7553), 436–444.
Li, Q., Sun, X.Y., Wang, L.X., et al., 2016. The use of vegetation supply water index (VSWI) based on different vegetation indices in the spring drought monitoring in Henan Province. Crops 1 (1), 162–168 (in Chinese).
Liu, W.T., Kogan, F.N., 1996. Monitoring regional drought using the vegetation condition index. Int. J. Remote Sens. 17 (14), 2761–2782.
Lu, X., Wei, M., Tang, G., et al., 2018. Evaluation and correction of the TRMM 3B43V7 and GPM 3IMERGM satellite precipitation products by use of ground-based data over Xinjiang, China. Environ. Earth Sci. 77 (5), 209.
Quiring, S.M., Ganesh, S., 2010. Evaluating the utility of the Vegetation Condition Index (VCI) for monitoring meteorological drought in Texas. Agric. For. Meteorol. 150 (3), 330–339.
Ran, Y.H., Li, X., Lu, L., 2010. Evaluation of four remote sensing based land cover products over China. Int. J. Remote Sens. 31 (2), 391–401.
Ren, M., 2013. Analysis on spatial and temporal characteristics drought of yunnan province. Acta Ecol. Sin. 33 (6), 317–324.
Rhee, J.Y., Im, J.H., Carbone, G.J., 2010. Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. Remote Sens. Environ. 114 (12), 2875–2887.
Shen, R.P., Guo, J., Zhang, J.X., et al., 2017. Construction of a drought monitoring model using the random forest based remote sensing. Journal of Geo-information Science 19 (1), 125–133 (in Chinese).
Singh, R.P., Roy, S., Kogan, F.N., 2003. Vegetation and temperature condition indices from NOAA AVHRR data for drought monitoring over India. Int. J. Remote Sens. 24 (22), 4393–4402.
Vicente-serrano, S.M., Beguería, S., López moreno, J.I., 2010. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. J. Clim. 23 (7), 1696–1718.
Vicente-Serrano, S.M., 2006. Differences in spatial patterns of drought on different time scales: an analysis of the Iberian Peninsula. Water Resour. Manag. 20 (1), 37–60.
Wang, P., 2004. Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains, USA. Int. J. Remote Sens. 25 (1), 61–72.
Wang, Q., Shi, P., Lei, T., et al., 2015. The alleviating trend of drought in the Huang-Huai-Hai Plain of China based on the daily SPEI. Int. J. Climatol. 35 (13), 3760–3769.
Wang, X., Liu, C., Cong, P., et al., 2018. Agriculture drought monitoring using remote sensing based on enhanced temperature vegetation dryness index. J. Arid Land Resour. Environ. 32 (5), 165–170 (in Chinese).
Yin, J., Zhan, X., Hain, C.R., Liu, J., Anderson, M.C., 2018. A method for objectively integrating soil moisture satellite observations and model simulations toward a blended drought index. Water Resour. Res. 54. https://doi.org/10.1029/2017WR021959.
Zhang, D., Zhang, W., Huang, W., et al., 2017. Upscaling of surface soil moisture using a deep learning model with VIIRS RDR. ISPRS Int. J. Geoinf. 6 (5), 130.